Almost Periodicity, Finite Automata Mappings and Related Effectiveness Issues

Yuri Pritykin*

February 7, 2008

Abstract

The paper studies different variants of almost periodicity notion. We introduce the class of eventually strongly almost periodic sequences where some suffix is strongly almost periodic (=uniformly recurrent). The class of almost periodic sequences includes the class of eventually strongly almost periodic sequences, and we prove this inclusion to be strict. We prove that the class of eventually strongly almost periodic sequences is closed under finite automata mappings and finite transducers. Moreover, an effective form of this result is presented. Finally we consider some algorithmic questions concerning almost periodicity.

1 Introduction

Strongly almost periodic sequences (=uniformly recurrent infinite words) were studied in the works of Morse and Hedlund [4, 5] and of many others (e. g., see [2, 7]). A sequence is strongly almost periodic if every its factor occurs infinitely many times with bounded distances. This notion first appeared in the field of symbolic dynamics, but then turned out to be interesting in connection with computer science, mathematical logic, combinatorics on words. Almost periodic sequences were introduced in [11] while studying logical theories of unary functions over $\mathbb{N}$. A sequence is almost periodic if every its factor either occurs infinitely many times with bounded distances or occurs only finitely many times. We introduce a new class of sequences called eventually strongly almost periodic, where some suffix is strongly almost periodic. Then we study some properties of this class.

This paper is organized as follows.

In Section 2 we give the formal definitions of different generalizations of periodicity notion. The class of almost periodic sequences includes the class of eventually strongly almost periodic sequences. We prove this inclusion to be strict (Theorem 1).

Section 3 concerns automata mappings. Almost periodic sequences were studied in detail in [12, 7]. In particular, the authors prove that the class of almost periodic sequences is closed under finite automata mappings (=mappings done by synchronizing finite transducers). Evidently, the class of finite automata mappings of strongly almost periodic sequences contains the class of eventually strongly almost periodic sequences. The main result of the article (Theorem 3) states the equality of the classes. In other words, Theorem 3 says that finite automata preserve the property of eventual strong almost periodicity. Moreover, an effective variant of this theorem is proved (Theorem 4).

*Moscow State University, Russia, e-mail: `pritykin@lpcs.math.msu.su`, `yura@mccme.ru`. The work was partially supported by RFBR grants 06-01-00122, 05-01-02803, Scientific Schools grant 358.2003.1 and Kolmogorov grant of Institute of New Technologies.

Then we consider a generalized version of finite automaton, i. e., finite transducer, and prove the same statement for it.

In Section 4 we deal with some algorithmic questions connected with almost periodicity. Namely, we prove that some problems or properties do not have corresponding effective analogs (in contrast to Theorem 3 with effective version in Theorem 4). For instance, we prove that given eventually strongly almost periodic sequence and its regulator we can not find any prefix which is enough to cut to obtain strongly almost periodic sequence.

Let us introduce some basic notions and notations.

Denote $\{0, 1\}$ by $\mathbb{B}$, the set of nonnegative integers $\{0, 1, 2, \dots\}$ by $\mathbb{N}$. Let Σ be a finite alphabet with at least two symbols. We consider the sequences over this alphabet, i. e., the mappings $\omega\colon \mathbb{N} \to \Sigma$. The set of all such sequences forms Cantor metric space. Denote this space by $\Sigma^{\mathbb{N}}$. Then $\lim_{n\to\infty} x_n = \omega$, if $\forall i\ \exists n\ \forall m > n\ x_m(i) = \omega(i)$ (this definition works for finite x_n too).

Denote by Σ^* the set of all finite strings over Σ including the empty string Λ. If $i \leq j$ are nonnegative integers, denote by $[i, j]$ the segment of $\mathbb{N}$ with ends in i and j, i. e., the set $\{i, i + 1, i + 2, \dots, j\}$. Also denote by $\omega[i, j]$ a substring $\omega(i)\omega(i + 1)\dots\omega(j)$ of a sequence ω. A segment $[i, j]$ is an occurrence of a string $x \in \Sigma^*$ in a sequence ω if $\omega[i, j] = x$. We say that $x \neq \Lambda$ is a factor of ω if x occurs in ω. The string of the form $\omega[0, i]$ for some i is called a prefix of ω, and respectively the sequence of the form $\omega(i)\omega(i + 1)\omega(i + 2)\dots$ for some i is called a suffix of ω and is denoted by $\omega[i, \infty)$. Denote by $|x|$ the length of a string x. The occurrence $x = \omega[i, j]$ in ω is k-aligned if $k|i$. Imagining the sequences going horizontally from the left to the right, we use terms "to the right" and "to the left" to talk about greater and smaller indices respectively.

2 Almost periodicity

A sequence ω is periodic if for some T we have $\omega(i) = \omega(i + T)$ for each $i \in \mathbb{N}$. This T is called a period of ω. The class of all periodic sequences we denote by $\mathcal{P}$. Let us consider some extensions of this class.

A sequence ω is called *almost periodic* if for any factor x of ω occurring in it infinitely many times there exists a number l such that any factor of ω of length l contains at least one occurrence of x. We denote the class of all almost periodic sequences by $\mathcal{AP}$.

A sequence ω is called *strongly almost periodic* if for any factor x of ω there exists a number l such that any factor of ω of length l contains at least one occurrence of x (and therefore x occurs in ω infinitely many times). Obviously, to show strong almost periodicity of a sequence it is sufficient to check the mentioned condition only for all prefixes but not for all factors. Denote by $\mathcal{SAP}$ the class of all strongly almost periodic sequences.

We also introduce an additional definition: a sequence ω is *eventually strongly almost periodic* if some its suffix is strongly almost periodic. The class of all eventually strongly almost periodic sequences we denote by $\mathcal{EAP}$.

Suppose $\omega \in \mathcal{EAP}$. Denote by $\mathrm{pr}(\omega)$ the minimal n such that $\omega[n, \infty) \in \mathcal{SAP}$. Thus for each $m \geqslant \mathrm{pr}(\omega)$ we have $\omega[m, \infty) \in \mathcal{SAP}$.

A function $\mathrm{R}_\omega\colon \mathbb{N} \to \mathbb{N}$ is an *almost periodicity regulator* of a sequence $\omega \in \mathcal{AP}$, if

(1) every string of length n occurring in ω infinitely many times, occurs on any factor of length $\mathrm{R}_\omega(n)$ in ω;
(2) any string of length n occurring finitely many times in ω, does not occur in $\omega[\mathrm{R}_\omega(n), \infty)$.

The latter condition is important only for sequences in $\mathcal{AP} \setminus \mathcal{SAP}$. Notice that regulator is not unique: any function greater than regulator is also a regulator.

Obviously, $\mathcal{P} \subset \mathcal{SAP} \subset \mathcal{EAP} \subset \mathcal{AP}$. In fact, all these inclusions are strict. For instance the famous Thue–Morse sequence $\omega_T = 0110100110010110\ldots$ (see [13, 1] or Section 4) is an example of the element in $\mathcal{SAP}$ but not in $\mathcal{P}$ (moreover, $\mathcal{SAP}$ has cardinality continuum while $\mathcal{P}$ is countable, see [3] or [7] for proofs). The inequality $\mathcal{SAP} \subsetneq \mathcal{EAP}$ is obvious. Let us prove $\mathcal{EAP} \subsetneq \mathcal{AP}$.

Theorem 1. *There exists a binary sequence $\omega \in \mathcal{AP} \setminus \mathcal{EAP}$.*

Proof. Construct a sequence of binary strings $a_0 = 1$, $a_1 = 10011$, $a_2 = 1001101100011001001110011$, and so on, by this rule:

$$a_{n+1} = a_n \overline{a}_n \overline{a}_n a_n a_n,$$

where $\overline{x}$ is a string obtained from x by changing every 0 to 1 and vice versa. Put

$$c_n = a_n a_n a_n a_n$$

and

$$\omega = c_0 c_1 c_2 c_3 \ldots$$

Let us prove that $\omega \in \mathcal{AP} \setminus \mathcal{EAP}$.

The length of a_n is 5^n, so the length of $c_0 c_1 \ldots c_{n-1}$ is $4(1 + 5 + \cdots + 5^{n-1}) = 5^n - 1$. By definition, put

$$l_n = 5^n - 1 = |c_0 c_1 \ldots c_{n-1}|.$$

Let us show that ω is almost periodic. Suppose $x \neq \Lambda$ occurs in ω infinitely many times. Take n such that $|x| < 5^n$. Suppose $[i, j]$ is an occurrence of x in ω such that $i \geq l_n$. By construction, for any k we can consider $\omega[l_k, \infty)$ as a concatenation of strings a_k and $\overline{a}_k$. Thus (by assumption about i) the string x is a substring of either $a_n a_n$, $a_n \overline{a}_n$, $\overline{a}_n a_n$ or $\overline{a}_n \overline{a}_n$. Notice that 10011 contains all strings of length two (00, 01, 10 and 11), so a_{n+1} contains each of $a_n a_n$, $a_n \overline{a}_n$, $\overline{a}_n a_n$, $\overline{a}_n \overline{a}_n$. Hence x is a substring of a_{n+1}. Similarly, x is a substring of $\overline{a}_{n+1}$. In each factor of length $2|a_{n+1}|$ of $\omega[l_{n+1}, \infty)$, a_{n+1} or $\overline{a}_{n+1}$ occurs. Hence for $l = (5^{n+1} - 1) + 2 \cdot 5^{n+1}$ the string x occurs on every factor of length l in ω.

Now let us prove that for any $n > 0$ the string c_n does not occur in $\omega[l_{n+1}, \infty)$. This implies that c_n occurs finitely many times in the suffix $\omega[l_n, \infty)$, i. e., this suffix is not strongly almost periodic. Therefore ω is not eventually strongly almost periodic.

Let $\nu = \omega[l_{n+1}, \infty)$. As above, for each k, $1 \leq k \leq n+1$, ν is a concatenation of strings a_k and $\overline{a}_k$. Assume c_n occurs in ν and let $[i, j]$ be one of this occurrences. For $n > 0$ the string c_n begins with a_1, hence $[i, i+4]$ is an occurrence of a_1 in ν. We see that $a_1 = 10011$ occurs in $a_1 a_1 = 1001110011$, $a_1 \overline{a}_1 = 1001101100$, $\overline{a}_1 a_1 = 0110010011$ or $\overline{a}_1 \overline{a}_1 = 0110001100$ only in 0th or 5th position. Thus $[i, j]$ is 5-aligned, hence ν and c_n can be considered as constructed of "letters" a_1 and $\overline{a}_1$, and we assume that c_n occurs in ν. Now it is easy to prove by induction on m that $[i, j]$ is 5^m-aligned for $1 \leqslant m \leqslant n$, i. e., we can consider ν and c_n to be constructed of "letters" a_m and $\overline{a}_m$, and assume that c_n occurs in ν (the base for $m = 1$ is already proved, and we can repeat the same argument changing 1 and 0 to a_m and $\overline{a}_m$ and taking into account that c_n begins with a_m for each $1 \leq m \leq n$).

Therefore we have shown that $[i, j]$ is 5^n-aligned, hence if we consider ν and c_n to be constructed of "letters" a_n and $\overline{a}_n$, then $c_n = a_n a_n a_n a_n$ occurs in ν. But notice that in any sequence constructed by concatenation of strings $a_1 = 10011$ and $\overline{a}_1 = 01100$ there is no any occurrence of 0000 or 1111. That is why c_n also can not occur in ν. This is a contradiction. $\square$

Moreover, it is quite easy to modify the proof and to show that $\mathcal{AP} \setminus \mathcal{EAP}$ has cardinality continuum. For instance for each sequence $\tau\colon \mathbb{N} \to \{4, 5\}$ we can construct ω_τ in the same way as in the proof of Theorem 1, but instead of c_n we take

$$c_n^{(\tau)} = \underbrace{a_n a_n \ldots a_n}_{\tau(n)}.$$

Obviously, all ω_τ are different for different τ and hence there exists continuum of various τ.

3 Finite automata mappings

It seems interesting to understand whether some transformations of sequences preserve the property of almost periodicity. The simplest type of algorithmic transformation is finite automaton mapping. Another motivation, less philosophical, is that finite automata mappings were one of the most useful tools in [12] while studying almost periodicity and finding some criterion for first-order and monadic theories of unary functions over $\mathbb{N}$ to be decidable.

Finite automaton is a tuple $F = \langle \Sigma, \Delta, Q, \tilde{q}, f \rangle$, where Σ and Δ are finite sets called input and output alphabets respectively, Q is a finite set of states, $\tilde{q} \in Q$ is the initial state, and

$$f\colon Q \times \Sigma \to Q \times \Delta$$

is the transition function. For $\alpha \in \Sigma^{\mathbb{N}}$ consider the sequence $\langle p_n, \beta(n) \rangle_{n=0}^{\infty}$, where $p_n \in Q$, $\beta(n) \in \Delta$, and assume $p_0 = \tilde{q}$ and $\langle p_{n+1}, \beta(n) \rangle = f(p_n, \alpha(n))$ for each n. Then we call $\beta = F(\alpha)$ a finite automata mapping of α. If $[i, j]$ is an occurrence of a string x in α, and $p_i = q$, then we say that automaton F comes to this occurrence of x being in the state q.

In [12, 7] the following statement was proved.

Theorem 2. *If F is a finite automaton and $\omega \in \mathcal{AP}$, then $F(\omega) \in \mathcal{AP}$.*

We can prove a counterpart of this statement for eventually strongly almost periodic sequences.

Theorem 3. *If F is a finite automaton and $\omega \in \mathcal{EAP}$, then $F(\omega) \in \mathcal{EAP}$.*

Proof. Obviously, it is enough to prove the theorem for $\omega \in \mathcal{SAP}$, since prefix does not matter.

Thus let $\omega \in \mathcal{SAP}$. By Theorem 2, $F(\omega) \in \mathcal{AP}$. Suppose $F(\omega)$ is not eventually strongly almost periodic. It means that for any natural N there exists a string that occurs in $F(\omega)$ after position N and does not occur after that. Indeed, if we remove the prefix $[0, N]$ from $F(\omega)$, we do not get strongly almost periodic sequence, hence there exists a string occurring in this sequence only finitely many times. Then take its rightmost occurrence.

Let $[i_0, j_0]$ be the rightmost occurrence of a string y_0 in $F(\omega)$. For some l_0 the string $x_0 = \omega[i_0, j_0]$ occurs in every factor of the length l_0 in ω (by the property of strong almost periodicity). If F comes to i_0 in the state q_0, then F never comes to righter occurrences of x in the state q_0 because in this case automaton outputs y_0 completely.

Now let $[r, s]$ be the rightmost occurrence of some string a in $F(\omega)$, where $r > i_0 + l_0$. On the factor $\omega[r - l_0, r]$ there exists an occurrence $[r', s']$ of the string x_0. By definition of r we have $r' > i_0$. Thus assume

$$i_1 = r',\ j_1 = s,\ x_1 = \omega[i_1, j_1],\ y_1 = F(\omega)[i_1, j_1].$$

Since a does not occur in $F(\omega)$ to the right of r, then y_1 does not occur in $F(\omega)$ to the right of i_1, for it contains a as a substring. Therefore if the automaton comes to the position i_1 in the state q_1,

then it never comes to righter occurrences of x_1 in the state q_1. Since x_1 begins with $\omega[r', s'] = x_0$, and $r' > i_0$, we get $q_1 \neq q_0$. We have found the string x_1 such that automaton F never comes to occurrences of x_1 to the right of i_1 in the state q_0 or q_1.

Let $m = |Q|$. Arguing as above, for $k < m$ we construct the strings $x_k = \omega[i_k, j_k]$ and corresponding different states q_k, such that F never comes to occurrences of x_k in ω to the right of i_k in the states $q_0, q_1, \ldots, q_k$. For $k = m$ we have a contradiction. □

Notice that this proof is non-effective in the following sense. Suppose we know $\omega \in \mathcal{SAP}$ and its almost periodicity regulator R_ω. Then by Theorem 3 some upper bound on $\mathrm{pr}(F(\omega))$ exists for $F(\omega) \in \mathcal{EAP}$, but the presented proof does not allow us to obtain any such bound.

Theorem 3 was proved first in [8]. The following effective version of this theorem was announced in [9].

For a function g denote $\underbrace{g \circ g \circ \cdots \circ g}_{n}$ by g^n.

Theorem 4. *Let F be a finite automaton with n states and $\omega \in \mathcal{SAP}$. Then $F(\omega) \in \mathcal{EAP}$ and*

$$\mathrm{pr}(F(\omega)) \leqslant \mathrm{R}_\omega^n(1) + \mathrm{R}_\omega^{n-1}(1) + \cdots + \mathrm{R}_\omega(1).$$

To prove this theorem, first we consider particular type of automata called reversible for which the statement of theorem is simple. Then we introduce some construction in combinatorics on words which allows us to reduce general situation to the case of reversible automaton.

A finite automaton $F = \langle \Sigma, \Delta, Q, \tilde{q}, f \rangle$ is *reversible*, if for any $q \in Q$ and $a \in \Sigma$ there exist unique $q' \in Q$ and $b \in \Delta$, such that $f(q', a) = \langle q, b \rangle$. In other words, in such an automaton each letter of the input alphabet Σ performs a permutation on Q (output alphabet does not matter). If we have some state, we can reconstruct the sequence of previous states from the sequence of previous input letters (this is a reversibility property).

Theorem 5. *If F is a reversible finite automaton and $\omega \in \mathcal{SAP}$, then $F(\omega) \in \mathcal{SAP}$.*

Proof. Suppose x occurs in ω, and F comes to this occurrence in the state q. Our goal is to prove that the next time when F comes to x in ω being in the state q, is at some distance of the previous such situation, and we can also give upper bound for this distance in terms of $|x|$ and R_ω. It means that the same estimate for this distance works for any situation when F comes to x in the state q. So it is enough for our purpose.

Let $x_0 = \omega[r, s]$ be a factor of ω; the automaton comes to this occurrence in some state q. Let $[i_0, j_0]$ be the next occurrence of x_0 in ω, then $j_0 \leqslant r + \mathrm{R}_\omega(|x|) + 1$. If F comes to this occurrence being in the state q, then all is done. Otherwise F comes to i_0 being in the state $q_0 \neq q$. Let $x_1 = \omega[r, j_0]$, and suppose $[i_1, j_1]$ is the next occurrence of x_1 in ω, so $j_1 \leqslant r + \mathrm{R}_\omega(\mathrm{R}_\omega(|x|) + 1) + 1$. Suppose the automaton comes to the position $i_1 + i_0$ being in the state q_1. If $q_1 = q$, then all is done for $\omega[i_1 + i_0, j_1] = x_0$. If $q_1 = q_0$, then F comes to the position i_1 in the state q, since F is reversible, and all is done again. If it is not the case, $q_1 \neq q$ and $q_1 \neq q_0$.

Similarly, for $x_2 = \omega[r, j_1]$ and its occurrence $[i_2, j_2]$ in ω, such that $i_2 > r$ and $j_2 \leqslant r + \mathrm{R}_\omega(\mathrm{R}_\omega(\mathrm{R}_\omega(|x|) + 1) + 1) + 1$, either we are done or F comes to the position $i_2 + i_1 + i_0$ in the state q_2 where $q_2 \neq q$, $q_2 \neq q_0$ and $q_2 \neq q_1$. Arguing in the same way, for $k < m = |Q|$ we construct the strings $x_0, x_1, \ldots, x_k$ with occurrences $[i_0, j_0], [i_1, j_1], \ldots, [i_k, j_k]$ and different states $q_0, q_1, \ldots, q_{k-1}$, such that in worst case F can not come to the position $i_k + i_{k-1} + \cdots + i_0$ in states $q, q_0, \ldots, q_{k-1}$. Thus for $k = m$ we are done for sure, and the estimate for distance will be $f(f(\ldots(|x|)\ldots))$, where $f = \mathrm{R}_\omega + 1$ and the number of iterations is m. □

For $\omega \in \Sigma^{\mathbb{N}}$, $\nu \in \Delta^{\mathbb{N}}$ define $\omega \times \nu \in (\Sigma \times \Delta)^{\mathbb{N}}$ such that $(\omega \times \nu)(i) = \langle \omega(i), \nu(i) \rangle$.

Corollary 6. *If $\omega \in \mathcal{SAP}$ and $\nu \in \mathcal{P}$, then $\omega \times \nu \in \mathcal{SAP}$.*

Proof. Operation "$\times$" with periodic sequence can be simulated by cyclic finite automaton which is obviously reversible. $\square$

Remark 7. Let us formulate some open questions connected with Corollary 6. Let $\omega, \nu \in \mathcal{SAP}$. Then it is interesting to know what we can say about $\omega \times \nu$. It is not difficult to construct $\omega \times \nu \notin \mathcal{AP}$ or $\omega \times \nu \in \mathcal{EAP} \setminus \mathcal{SAP}$. Can we construct $\omega \times \nu \in \mathcal{AP} \setminus \mathcal{EAP}$? Does there exist any criterion to determine whether $\omega \times \nu \in \mathcal{AP}$? An example of $\omega, \nu \in \mathcal{AP}$ with $\omega \times \nu \notin \mathcal{AP}$ can be found in [7].

Now consider the following construction. Let $\omega \in \Sigma^{\mathbb{N}}$, and suppose $a \in \Sigma$ occurs in ω infinitely many times. Cut ω on the blocks like xa, where $x \in (\Sigma \setminus \{a\})^*$, i. e., on the blocks containing a symbol a on the end and not containing any other occurrences of a. To make this we need to cut after each occurrence of a. If a occurs in ω at bounded distance, then the number of all such blocks is finite (for example, if $\omega \in \mathcal{AP}$, then the length of blocks is not more then $\mathrm{R}_\omega(1)$). Encode these blocks by symbols of some finite alphabet, denote this alphabet by $\mathrm{b}_{a,\omega}(\Sigma)$. Thus we obtained a new sequence in this alphabet from ω. Delete the first symbol of this sequence. The result is called an *a-split* of ω and is denoted by $\mathrm{s}_a(\omega)$. For example, 0-split of the sequence 3200122403100110... is (0)(12240)(310)(0)(110)...

Lemma 8. *Let $\omega \in \mathcal{SAP}$, and suppose $a \in \Sigma$ occurs in ω. Then $\mathrm{s}_a(\omega) \in \mathcal{SAP}$.*

Proof. Let k be the maximal length of the a-split blocks. Consider a prefix x of $\mathrm{s}_a(\omega)$. The corresponding string y in ω is not longer than $k|x|$. Let $z = ay$, $|z| \leqslant k|x| + 1$. The string z occurs in ω. Therefore z occurs on any factor of length $l = \mathrm{R}_\omega(k|x| + 1)$ in ω. The first and the last symbols of z are a, so every such occurrence is well-aligned relative to a-split of ω. Hence for any occurrence of z in ω there is an occurrence of x in $\mathrm{s}_a(\omega)$. Therefore x occurs on each factor of length $\mathrm{R}_\omega(k|x| + 1)$ in $\mathrm{s}_a(\omega)$. $\square$

Remark 9. In connection with Lemma 8 an interesting question appears, a particular case of the question in Remark 7. Instead of splitting ω on blocks with some fixed symbol at the end, we can split ω arbitrarily on blocks of various lengths. When the result is strongly almost periodic or just almost periodic?

Now we can prove the promised theorem.

Proof of Theorem 4. Let $F = \langle \Sigma, \Delta, Q, \tilde{q}, f \rangle$ and $|Q| = n$. We construct an algorithm to compute some

$$l \geqslant \mathrm{pr}(F(\omega)),$$

and at the same time we prove

$$l \leqslant \mathrm{R}^n_\omega(1) + \mathrm{R}^{n-1}_\omega(1) + \cdots + \mathrm{R}_\omega(1).$$

Assume that any automaton in the proof has the maximum possible output alphabet "input alphabet"$\times$"the set of states" (general case can be obtained from this by projection). For example for F this is $\Sigma \times Q$. Correspondingly the transition function f writes the pair of a current state and an input symbol to the output. Further we omit the second part of the transition function value (i. e., for instance, instead of $f(p, a) = \langle q, b \rangle$ we write just $f(p, a) = q$ with $f(p, a) = \langle q, \langle p, a \rangle \rangle$ in mind).

Let $\omega_0 = \omega$. Suppose every symbol of Σ occurs in ω_0, elsewise we restrict F only on the symbols occurring in ω_0; to determine these symbols effectively we should read first $\mathrm{R}_{\omega_0}(1)$ symbols of ω_0.

If F is reversible, by Theorem 5 we get $\mathrm{pr}(F(\omega_0)) = 0$. Otherwise some symbol $a_0 \in \Sigma$ accomplishes not one-to-one mapping of Q, so the set

$$Q_1 = \{q \ : \ \exists q' \ \ f(q', a_0) = q\}$$

is the proper subset of Q. Consider

$$\omega_1 = \mathrm{s}_{a_0}(\omega_0),$$

which is strongly almost periodic by Lemma 8. Notice that starting in any state on ω_0, the automaton F comes to any block of a_0-split of ω_0 in the state of the set Q_1, because every such block has a_0 at the end.

Let us construct a new automaton F_1 (effectively by F). Let the input alphabet of F_1 be $\mathrm{b}_{a_0,\omega_0}(\Sigma)$, the set of states be Q_1, and the value of the transition function on $x \in \mathrm{b}_{a_0,\omega_0}(\Sigma)$, $q \in Q_1$ be the output of F starting in the state q on the string x written in symbols of Σ. Let the initial state of F_1 be the state of F after the work on the prefix of ω until the first occurrence of a_0 (the prefix which we delete obtaining $\mathrm{s}_{a_0}(\omega_0)$ from ω_0). Now the work of F_1 on ω_1 simulates the work of F on ω_0. Notice that ω_1 is obtained from ω by deleting not more than $\mathrm{R}_{\omega_0}(1)$ first symbols, counting in the alphabet Σ.

We have the sequence ω_1 (in the alphabet more than initial) and the automaton F_1 with the set of states less than initial. If F_1 is not reversible, then we can repeat the procedure of the last paragraph. Thus we obtain the sequence ω_2 in some alphabet $\mathrm{b}_{a_1,\omega_1}(\mathrm{b}_{a_0,\omega_0}(\Sigma))$, and the automaton F_2 with the set of states less than previous, working on ω_2. The sequence ω_2 is obtained from ω_1 by deleting not more than $\mathrm{R}_{\omega_1}(1)$ first symbols, counting in the alphabet $\mathrm{b}_{a_0,\omega_0}(\Sigma)$. Therefore ω_2, written in the initial alphabet Σ, is obtained from ω by deleting not more than $\mathrm{R}_{\omega_0}(\mathrm{R}_{\omega_0}(1))+\mathrm{R}_{\omega_0}(1)$ first symbols, counting in the alphabet Σ.

An automaton with a single state (and with arbitrary input alphabet) is always reversible. Hence after k repetitions of described procedure for some $k \leqslant n$, we get the situation when the reversible automaton F_k works on strongly almost periodic sequence ω_k in some alphabet (after each repetition of the procedure the number of states decreases). The symbols of this alphabet code the blocks of initial sequence. Thus $F_k(\omega_k) \in \mathcal{SAP}$.

Writing ω_k in the alphabet Σ, we get some suffix ω' obtained from ω by deleting a prefix not longer than

$$\mathrm{R}^k_\omega(1) + \mathrm{R}^{k-1}_\omega(1) + \cdots + \mathrm{R}_\omega(1) \leqslant \mathrm{R}^n_\omega(1) + \mathrm{R}^{n-1}_\omega(1) + \cdots + \mathrm{R}_\omega(1).$$

It only remains to check why $F_k(\omega_k) \in \mathcal{SAP}$ implies $F(\omega') \in \mathcal{SAP}$. Let us explain this in one simple case when the automaton F_1 obtained after the first iteration of procedure is reversible (the general situation can be reduced to this case by induction). Then ω' is obtained from ω by deleting first symbols until the first occurrence of a_0. Let the initial state of F_1 (in which F comes to ω') be q. To show $F(\omega') \in \mathcal{SAP}$ it is necessary and sufficient to check whether for any prefix of ω' the occurrences of its copies in ω', in which F comes in the state q, are quite regular, i. e. these copies occur on each factor of length l for some l (sufficient condition is obvious, necessary condition follows from our requirement for automata always to output the pair $\langle$input symbol, current state$\rangle$).

Let x be a prefix of ω' ending by a_0 (arbitrary prefix is contained in some such prefix). We can correctly split it on blocks ending by a_0, denote this split by y. The automaton F_1 is reversible, so $F_1(\omega_1) \in \mathcal{SAP}$. By the necessary and sufficient condition of the previous paragraph F_1 comes to y in the state q on each factor of length t for some t. Every such situation corresponds in ω' to coming F to some occurrence of x in the state q, and this happens on each factor of length tk, where $k \leqslant \mathrm{R}_\omega(1)$ is the maximal length of blocks. $\square$

The upper bound on $\mathrm{pr}(F(\omega))$ in Theorem 4 can not be significantly improved, as it follows from the construction in [10] after small modifications.

It is interesting that now we have two different proofs of Theorem 3, and it seems that there is no any connection between them.

The results about finite automata mappings can be extended on more general class of mappings done by finite transducers.

Let Σ and Δ be finite alphabets. The mapping $h\colon \Sigma^* \to \Delta^*$ is called a *homomorphism*, if for any $u, v \in \Sigma^*$ we have $h(uv) = h(u)h(v)$. Clearly, any homomorphism is fully determined by its values on single-letter strings. Let $\omega \in \Sigma^{\mathbb{N}}$. By definition, put

$$h(\omega) = h(\omega(0))h(\omega(1))h(\omega(2))\dots$$

Suppose $h\colon \Sigma^* \to \Delta^*$ is a homomorphism, $\omega \in \Sigma^{\mathbb{N}}$ is almost periodic. In [7] it was shown that if $h(\omega)$ is infinite, then it is almost periodic. Thus obviously if ω is strongly almost periodic, and $h(\omega)$ is infinite, then $h(\omega)$ is also strongly almost periodic. Indeed, it is enough to show that any v occurring in $h(\omega)$ occurs infinitely many times. However there exists some factor u of ω such that $h(u)$ contains v, but by the definition of strong almost periodicity u occurs in ω infinitely many times. Evidently, for $\omega \in \mathcal{EAP}$ we have $h(\omega) \in \mathcal{EAP}$, if $h(\omega)$ is infinite.

Now we modify the definition of finite automaton, allowing it to output any string (including the empty one) over output alphabet reading only one character from input. This modification is called *finite transducer* (see [14]). Formally, we only change the definition of translation function. Now it has the form

$$f\colon Q \times \Sigma \to Q \times \Delta^*.$$

If the sequence $\langle p_n, v_n\rangle_{n=0}^{\infty}$, where $p_n \in Q$, $v_n \in \Delta^*$, is the mapping of α, then the output is the sequence $v_0 v_1 v_2 \dots$

Actually, we can decompose the mapping done by finite transducer into two: the first one is a finite automaton mapping and another is a homomorphism. Each of these mappings preserves the class $\mathcal{AP}$, so we get the corollary: finite transducers map almost periodic sequences to almost periodic. Similarly, by Theorem 3 and arguments above we also get the following

Corollary 10. *Let F be a finite transducer, $\omega \in \mathcal{EAP}$. Suppose $F(\omega)$ is infinite. Then $F(\omega) \in \mathcal{EAP}$.*

4 Effectiveness

Lots of interesting algorithmic questions naturally appear in connection with almost periodicity, i. e., if one can check some property or find some characteristic algorithmically being given a sequence. Sometimes these questions are just effective issues for corresponding noneffective results, for example Theorem 4 is an effective variant of Theorem 3. Further, we mainly deal with the case when the answers on these questions are negative. We prove that some properties do not have effective analogs.

Formally, we consider an algorithm with an oracle for a sequence on input. This algorithm halts on every oracle and outputs a finite binary string or any other constructive object. The main property of such an algorithm is continuity: it outputs the answer on having read only finite number of symbols from the sequence. Thus to prove non-effectiveness we only need to show discontinuity. In fact, such proofs are just concrete complicated combinatorial constructions showing this discontinuity.

If we have only a sequence, then we can not recognize almost any property about this sequence. For example it is even impossible to understand whether the symbol 1 occurs in given binary sequence: if an algorithm checks some finite number of symbols and all these symbols are 0, then it can not guarantee that 1 does not occur further. The question about algorithmic decidability becomes more interesting if we allow to give on input some additional information. In the case of almost periodic sequences it may be an almost periodicity regulator.

It is easy to decode unambiguously functions $\mathbb{N} \to \mathbb{N}$ and also pairs $\langle$sequence, function$\rangle$ by binary sequences. That is why we can correctly consider algorithms with an almost periodic sequence ω and its regulator f on input.

From this point of view the above problem can be solved effectively: reading first $f(1)$ symbols of the sequence we can say whether or not 1 occurs in it, and moreover reading next $f(1)$ symbols we can say whether 1 occurs in it finitely or infinitely many times.

The following several theorems are examples of problems concerning almost periodicity which do not have effective analogs. It is especially interesting that the Theorem 11 results are absolutely contrary to the results of Theorem 4. Theorem 14 is also connected with Theorem 4. All the following theorems were announced first in [9].

We say $f_n \to f$ for $f_n, f \colon \mathbb{N} \to \mathbb{N}$ if $\forall i\ \exists n\ \forall m > n\ f_m(i) = f(i)$.

Theorem 11. *Given $\omega \in \mathcal{EAP}$ and its regulator f, it is impossible to compute algorithmically some $l \geqslant \mathrm{pr}(\omega)$.*

Remind that ω_T is the Thue–Morse sequence. This sequence can be obtained as follows: let $a_0 = 0$, $a_{n+1} = a_n\overline{a}_n$, and $\omega_T = \lim_{n\to\infty} a_n$. Notice that $|a_n| = 2^n$. The Thue–Morse sequence has lots of interesting properties (see [1]), but we are interested in the following one: ω_T is cube-free, i. e., for any $a \in \mathbb{B}^*$, $a \neq \Lambda$ the string aaa does not occur in ω_T (see [1, 13]).

Proof of Theorem 11. It is enough to construct $\omega_n \in \mathcal{EAP}$, $\omega \in \mathcal{SAP}$ with regulators f_n, f such that $\omega_n \to \omega$, $f_n \to f$, but $\mathrm{pr}(\omega_n) \to \infty$. Indeed, suppose the mentioned algorithm exists and it outputs some $l \geqslant 0$ (arbitrary for $\omega \in \mathcal{SAP}$) given $\langle \omega, f\rangle$ on the input. During the computation of l the algorithm reads only finite number of symbols in ω and of values of f. Hence there exists $N > l$ such that algorithm does not know any $\omega(k)$ or $f(k)$ for $k > N$. Since $\mathrm{pr}(\omega_n) \to \infty$, there exists n such that $\mathrm{pr}(\omega_n) > N$. The algorithm works on the input $\langle \omega_n, f_n\rangle$ in the same way as it works on the input $\langle \omega, f\rangle$, and then outputs l, but $\mathrm{pr}(\omega_n) > N > l$.

Let $\omega = \omega_T$, $\omega_n = a_na_na_n\omega$. Notice that $\mathrm{pr}(\omega_n) \geqslant 2^n$. Indeed, if $\mathrm{pr}(\omega_n) < 2^n$, then $a_na_n\omega = a_na_na_n\overline{a}_n\overline{a}_na_n\ldots \in \mathcal{SAP}$, and hence $a_na_na_n$ occurs in ω_T — contradiction with the statement before the proof.

It only remains to show that we can find regulators f_n, f for ω_n, ω such that $f_n \to f$. It is sufficient to find the same regulator g for all ω_n (then we can increase it and obtain the same regulator for all ω_n and for ω too). Fix some R_ω and assume $g = 4\cdot \mathrm{R}_\omega$. Let v, $|v| = k$ occur in $\omega_n = a_na_na_n\omega$ infinitely many times. Let us take the factor $\omega[i,j]$ of length $4\cdot \mathrm{R}_\omega(k)$ and show that v occurs in it. If $j \geqslant 3\cdot 2^n + \mathrm{R}_\omega(k)$, then v occurs on the factor $\omega[3\cdot 2^n, 3\cdot 2^n + \mathrm{R}_\omega(k)]$ (by definition of R_ω). Otherwise $j < 3\cdot 2^n + \mathrm{R}_\omega(k)$, hence $i \leqslant 3\cdot 2^n - 3\,\mathrm{R}_\omega(k)$. But $i \geqslant 0$, therefore $\mathrm{R}_\omega(k) \leqslant 2^n = |a_n|$. Then $\omega_n[i, i+\mathrm{R}_\omega(k)]$ is contained in a_na_n. But a_na_n occurs in ω, so $\omega_n[i, i+\mathrm{R}_\omega(k)]$ occurs too. Therefore v occurs in ω.

However g is not already required. We should watch on the strings occurring in ω_n finitely many times. Obviously, if some v occurs in ω_n finitely many times, then $|v| = k > 2^n$ (otherwise v occurs in two consequent strings a_n or $\overline{a}_n$, and thus in ω). Therefore this can happen only for finite number of different n. Considering all the situations when strings of length k occur in some

ω_n finitely many times, we probably increase the value $g(k)$, but only finitely many times. Thus the required estimate for regulators exists. $\square$

We have already seen that $\mathcal{EAP} \subsetneq \mathcal{AP}$ (Theorem 1). Using the same construction we can show that it is impossible to separate these classes effectively.

Theorem 12. *Given $\omega \in \mathcal{AP}$ and its regulator f, it is impossible to determine algorithmically whether $\omega \in \mathcal{EAP}$.*

In [7] the following universal method for construction of strongly almost periodic sequences was presented. This method is based on block algebra on words introduced in [6] and then studied in [3].

The sequence $\langle A_n, l_n \rangle$, where $A_n \subset \Sigma^*$ for some finite alphabet Σ, $l_n \in \mathbb{N}$, is called *strong Σ-scheme*, if the following conditions hold:

(1) all the strings in A_n have the length l_n;
(2) any string $u \in A_{n+1}$ has the form $u = v_1 v_2 \dots v_k$, where $v_i \in A_n$, and for every $w \in A_n$ there exists i such that $v_i = w$.

We say that $\alpha \in \Sigma^{\mathbb{N}}$ is generated by strong Σ-scheme $\langle A_n, l_n \rangle$ if for every i and n we have

$$\alpha[il_n, (i+1)l_n - 1] \in A_n.$$

It is easy to see (by compactness) that any strong scheme generates some sequence. In [7], the authors prove that any sequence generated by strong scheme is strongly almost periodic. Moreover every strongly almost periodic sequence is generated by some strong scheme.

Proof of Theorem 12. It is enough to construct $\omega_n \in \mathcal{EAP}$, $\omega \in \mathcal{AP} \setminus \mathcal{EAP}$ with the same regulator f for all ω_n such that $\omega_n \to \omega$.

In the same way as in Theorem 1, assume $a_0 = 1$, and then by the rule: $a_{n+1} = a_n \overline{a}_n \overline{a}_n a_n a_n$. Denote $a_n a_n a_n a_n$ by c_n. Put $l_n = 5^n - 1 = |{}_0c_1 \dots {}_{n-1}|$. Consider $\omega = {}_{0123}\dots$ and $\nu = \lim_{n\to\infty} a_n$. From the proof of Theorem 1 it follows that $\omega \in \mathcal{AP} \setminus \mathcal{EAP}$. Let $\omega_n = c_0 c_1 \dots c_n \nu$. The sequence ν is generated by the strong $\mathbb{B}$-scheme $\langle \{a_n, \overline{a}_n\}, 5^n \rangle$, hence $\nu \in \mathcal{SAP}$. Therefore $\omega_n \in \mathcal{EAP}$. Obviously, $\omega_n \to \omega$, and it only remains to find common regulator f. We will get finite number of conditions of the form $f(k) \geqslant \alpha$, then we can take the maximum among all these α.

Let $v = \omega_n[i,j]$, $|v| = k$ occurs in $\omega_n = c_0 c_1 \dots c_n \nu$ infinitely many times. Then v occurs in ν, hence in a_m for some m too. Therefore v occurs in ω infinitely many times and it is sufficient to take $f(k) \geqslant \mathrm{R}_\omega(k) + \mathrm{R}_\nu(k)$.

Let $v = \omega_n[i,j]$, $|v| = k$ occurs in ω_n finitely many times. Then $i < l_n$. Suppose $j > l_n$. If $k \leqslant 5^n$, then v occurs in a_m for some m and hence occurs in ν infinitely many times. But the inequality $k > 5^n$ holds only for finite number of different n, and this yields just finitely many conditions on $f(k)$. Now suppose $j \leqslant l_n$. But then v occurs in $c_0 c_1 \dots c_n$ and occurs in ω finitely many times (otherwise v occurs in a_m for some m). Therefore in this case it is sufficient to take $f(k) \geqslant \mathrm{R}_\omega(k)$. $\square$

The following theorem shows that it is even impossible to separate effectively $\mathcal{SAP}$ and $\mathcal{P}$.

Theorem 13. *Given $\omega \in \mathcal{SAP}$ and its regulator f, it is impossible to determine algorithmically whether $\omega \in \mathcal{P}$.*

Proof. It is enough to construct $\omega_n \in \mathcal{P}$, $\omega \in \mathcal{SAP} \setminus \mathcal{P}$ with common regulator f for all ω_n such that $\omega_n \to \omega$.

Every strongly almost periodic sequence can be obtained from the strong Σ-scheme $\langle A_n, l_n\rangle$. Let us strengthen the main condition on A_n: let us consider strong schemes such that for each $n \in \mathbb{N}$ every $u \in A_{n+1}$ has the form $u = v_1 v_2 \dots v_k$, where $v_i \in A_n$, and for any $w_1, w_2 \in A_n$ there exists $i < k$ such that $v_i v_{i+1} = w_1 w_2$. Notice that such schemes exist and can generate non-periodic sequences, e. g., $\langle \{a_n, \overline{a}_n\}, 2^n\rangle$ generates ω_T.

Let $\langle A_n, l_n\rangle$ be the strong scheme satisfying the strengthened condition from the previous paragraph, generating $\omega \notin \mathcal{P}$. Let $p_n = \omega[0, l_n]$. Thus $p_n \in A_n$ and $\lim_{n\to\infty} p_n = \omega$. Assume $\omega_n = p_n p_n p_n \dots \in \mathcal{P}$. Obviously $\omega_n \to \omega$ and it only remains to find some common regulator f for all ω_n.

Let $v = \omega_n[i, j]$, $|v| = k$ (since $\omega_n \in \mathcal{P}$, it follows that v occurs in ω_n infinitely many times). The inequality $k \geqslant |p_n| = l_n$ holds only for finite number of different n, and this yields just finitely many conditions on $f(k)$. Now we can assume that $k < l_n$. Take t such that $l_{t-1} < k \leqslant l_t$ (it is important that t does not depends on n and is uniquely defined by k). Then $t < n$. There exists m such that $ml_t \leqslant i$ and $j \leqslant (m+2)l_t$, i. e., v occurs in some ab, where $a, b \in A_t$. Then by the scheme property v occurs in any $c \in A_{t+1}$. But on every factor of ω_n of length $2l_{t+1}$ there exists an occurrence of some $c \in A_{t+1}$ (fully contained in some p_n). Therefore it is sufficient to take $f(k) \geqslant 2l_{t+1}$. □

By the argument of Theorem 13 we obtain that there exists infinite set of periodic sequences with common regulator (while the period tends to infinity). This construction can be used in the following theorem: adding one symbol to the strongly almost periodic sequence we can not check whether it is still strongly almost periodic.

Theorem 14. *Given $\omega \in \mathcal{EAP}$, its regulator f and some $l \geqslant \mathrm{pr}(\omega)$, it is impossible to find algorithmically* $\mathrm{pr}(\omega)$.

Lemma 15. *If $a\omega \in \mathcal{SAP}$ for $a \in \Sigma^*$ and $\omega \in \mathcal{P}$ with period l, then $a\omega \in \mathcal{P}$ with period l.*

Proof. It is enough to prove lemma for single-letter a. Let $\alpha = 012\dots(l-1)012\dots(l-1)012\dots(l-1)\dots$ be periodic sequence over alphabet $\Sigma_l = \{0, 1, 2, \dots, l-1\}$. Then by Corollary 6 we have $a\omega \times \alpha \in \mathcal{SAP}$. In this sequence the symbol $\langle a, 0\rangle$ occurs infinitely many times, hence $a = \omega(l)$. □

Proof of Theorem 14. It is enough to construct $\omega_n \in \mathcal{EAP}$, $\omega \in \mathcal{SAP}$ with common regulator f for all ω_n such that $\omega_n \to \omega$ and $\mathrm{pr}(\omega_n) = 1$ ($\omega \in \mathcal{SAP}$ means $\mathrm{pr}(\omega) = 0$).

Notice that $1\omega_T \in \mathcal{SAP}$. Indeed, for each n strings $a_n a_n$ and $\overline{a}_n a_n$ occur in ω_T, and hence $1a_n$ too. Analogously $0\omega_T \in \mathcal{SAP}$.

By proof of Theorem 13, we can choose sequence $k_n \to \infty$ such that all periodic sequences like $\omega(0)\dots\omega(k_n)\omega(0)\dots\omega(k_n)\omega(0)\dots$ have common regulator f. Take a subsequence m_n of the sequence k_n such that all the symbols $\omega(m_n)$ are equal. Suppose these symbols are 0.

Let $\omega_n = 1\omega(0)\dots\omega(m_n)\omega(0)\dots\omega(m_n)\omega(0)\dots$ and $\omega = 1\omega_T$. There exists common estimate g on the regulator for these sequences. Indeed, it is sufficient $g(k) \geqslant f(k) + 1$ (by considering strings occurring infinitely many times) and $g(k) \geqslant k$ (by considering strings occurring only finitely many times: this can happen only for prefixes occurring exactly once).

If $\omega_n \in \mathcal{SAP}$, then by Lemma 15 we have $\omega_n \in \mathcal{P}$ with period m_n. But $\omega_n(0) = 1 \neq \omega_n(m_n) = 0$. Therefore $\mathrm{pr}(\omega_n) = 1$.

The case when all the symbols $\omega(m_n)$ are 1 is analogous (then ω_n begins with 0). □

5 Acknowledgements

The author is grateful to A. Semenov and An. Muchnik for their help in the work and also to M. Raskin, A. Rumyantsev, A. Shen and to all other participants of Kolmogorov seminar (Moscow) for useful discussions. The results were also presented on the seminar under the direction of S. Adian, the author is thankful to all the participants for attention.

References

[1] J.-P. Allouche, J. Shallit. *The ubiquitous Prouhet–Thue–Morse sequence.* Sequences and their applications, Proceedings of SETA'98, Springer Verlag, pp. 1–16, 1999.

[2] J. Cassaigne. *Recurrence in infinite words.* Proceedings of the 18th Symposium on Theoretical Aspects of Computer Science (STACS 2001), Springer Verlag, pp. 1–11, 2001.

[3] K. Jacobs. *Maschinenerzeugte 0-1-Folgen.* Selecta Mathematica II. Springer Verlag: Berlin, Heidelberg, New York, 1970.

[4] M. Morse, G. A. Hedlund. *Symbolic dynamics.* American Journal of Mathematics, 60, pp. 815–866, 1938.

[5] M. Morse, G. A. Hedlund. *Symbolic dynamics II: Sturmian trajectories.* American Journal of Mathematics, 62, pp. 1–42, 1940.

[6] M. Keane. *Generalized Morse sequences.* Z. Wahrsheinlichkeitstheorie verw. Geb., 10, pp. 335–353, 1968.

[7] An. Muchnik, A. Semenov, M. Ushakov. *Almost periodic sequences.* Theoretical Computer Science, vol. 304, pp. 1–33, 2003.

[8] Yu. L. Pritykin. *Strongly Almost Periodic Sequences under Finite Automata Mappings.* Matematicheskie Zametki, to appear (in Russian), 2006. English version on `http://arXiv.org/abs/cs/0605026`.

[9] Yu. L. Pritykin. *Finite automata mappings of strongly almost periodic sequences and algorithmic undecidability.* Proceedings of XXVIII Conference of Young Scientists, Moscow State University, Faculty of Mechanics and Mathematics, to appear, 2006 (in Russian).

[10] M. A. Raskin. *On the estimate of the regulator for automaton mapping of almost periodic sequence.* Proceedings of XXVIII Conference of Young Scientists, Moscow State University, Faculty of Mechanics and Mathematics, to appear, 2006 (in Russian).

[11] A. L. Semenov. *On certain extensions of the arithmetic of addition of natural numbers.* Math. of USSR, Izvestia, vol. 15, pp. 401–418, 1980.

[12] A. L. Semenov. *Logical theories of one-place functions on the set of natural numbers.* Math. of USSR, Izvestia, vol. 22, pp. 587–618, 1983.

[13] A. Thue. *Über unendliche Zeichenreihen.* Norske vid. Selsk. Skr. Mat. Nat. Kl., 7, pp. 1–22, 1906. Reprinted in *Selected mathematical papers of Axel Thue*, Universitetsforlaget, Oslo, pp. 139–158, 1977.

[14] A. Weber. *On the valuedness of finite transducers.* Acta Informatica, 27, pp. 749–780, 1989.